DIFFERENCES IN EARNINGS ACROSS SEX AND EDUCATIONAL ATTAINMENT: COMPARING BOX PLOTS

TEACHER VERSION

Subject Level:
High School Math

Grade Level:
9

Approx. Time Required:
90 minutes

Learning Objectives:

- Students will be able to interpret a box plot.

- Students will be able to compare different variables across data sets.

- Students will be able to compare box plots to make claims and inferences based on the data.

United States Census Bureau

STATISTICS IN SCHOOLS

Activity Description

Students will interpret box plots that represent the national median earnings of men and women aged 25 and older whose highest levels of educational attainment are either a high school diploma (or equivalent) or a bachelor's degree. Students will use the box plots to identify each data set's median, maximum, minimum, first quartile, third quartile, range, interquartile range, and outliers. They will also compare the box plots to draw conclusions about differences in earnings between the sexes and between levels of educational attainment.

Suggested Grade Level:
 9

Approximate Time Required:
 90 minutes

Learning Objectives:
- Students will be able to interpret a box plot.
- Students will be able to compare different variables across data sets.
- Students will be able to compare box plots to make claims and inferences based on the data.

Topics:
- Box plots
- Interquartile range
- Levels of educational attainment
- Median
- Median earnings
- Outliers
- Quartiles

Skills Taught:
- Comparing and interpreting box plots

Materials Required

- The student version of this activity, 8 pages
- Calculators

Activity Item

The following item is part of this activity. The item, its data source, and instructions for viewing the source data online appear at the end of this teacher version.

- Item 1: Box Plots

For more information to help you introduce your students to the U.S. Census Bureau, read "*Census Bureau 101 for Students*." This information sheet can be printed and passed out to your students as well.

Standards Addressed

See charts below. For more information, read "*Overview of Education Standards and Guidelines Addressed in Statistics in Schools Activities*."

Common Core State Standards for Mathematics

Standard	Domain	Cluster
CCSS.MATH.CONTENT.HSS.ID.A.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.	ID – Interpreting Categorical & Quantitative Data	Summarize, represent, and interpret data on a single count or measurement variable.
CCSS.MATH.CONTENT.HSS.ID.A.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).	ID – Interpreting Categorical & Quantitative Data	Summarize, represent, and interpret data on a single count or measurement variable.

Common Core State Standards for Mathematical Practice

Standard
CCSS.MATH.PRACTICE.MP2. Reason abstractly and quantitatively.   Students will make sense of quantities and their relationships in context. Specifically, students will interpret the data to determine the meaning of differences in median earnings between the sexes.
CCSS.MATH.PRACTICE.MP3. Construct viable arguments and critique the reasoning of others.   Students will make plausible arguments that consider the context of the data. Students will also respond to other students' arguments as part of a class discussion.
CCSS.MATH.PRACTICE.MP4. Model with mathematics.   Students will analyze relationships evident in the data to draw conclusions, interpreting their mathematical results in a real-world context.

National Council of Teachers of Mathematics' Principles and Standards for School Mathematics

Content Standard	Students should be able to:	Expectation for Grade Band
Data Analysis and Probability	Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.	Understand histograms, parallel box plots, and scatterplots and use them to display data.

Guidelines for Assessment and Instruction in Statistics Education

GAISE	Level A	Level B	Level C
Formulate Questions	X		
Collect Data			
Analyze Data		X	
Interpret Results		X	

Bloom's Taxonomy

Students will *analyze* box plots to determine values present in the data, to compare variables, and to *evaluate* those variables' relationships to each other.

Creating
Evaluating
Analyzing
Applying
Understanding
Remembering
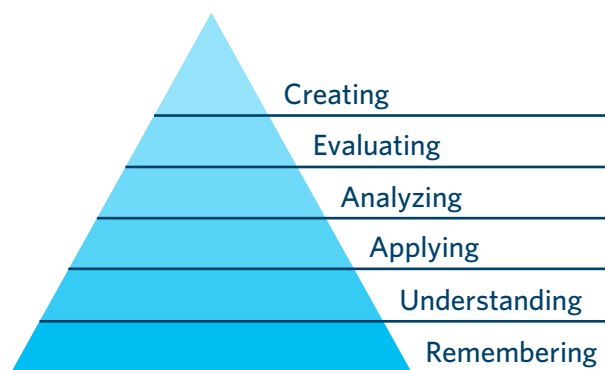
Teacher Notes

Before the Activity

Students must understand the following key terms:

- **Box plot** – a method of visually displaying a data set using the median, quartiles, and extremes of the data set

- **Measure of variability** – a measure of how spread out the values are in a data set

- **Mean** – a measure of center in a set of numerical data, computed by adding the values in a list and then dividing by the number of values in the list

- **Median** – a measure of center in a set of numerical data, identified as the value appearing at the middle of a sorted version of the list (or the mean of the two middle values if the list contains an even number of values)

- **Interquartile range (IQR)** – a measure of variability in a set of numerical data to indicate the difference between the first and third quartiles of the data set

- **Outlier** – an extremely high or low value that noticeably differs from the other data points in the set; it is any data point that falls 1.5 times the IQR below the first quartile or above the third quartile.

Students should have the following skills:

- Ability to identify the median, quartiles, minimum, maximum, range, interquartile range, 5-number summary, and outliers from a box plot

- Ability to compute the median, range, and interquartile range

- Teachers should review the sample answers to see teacher notes on where in the activity students might need additional guidance. Teachers could preview the activity with students to draw their attention to the italicized terms in questions 7 and 8 for enhanced understanding. Teachers may also want to review quartiles so that students can confidently answer these two questions.

The teacher should explain to students that the data in this activity come from the American Community Survey, a survey conducted monthly by the Census Bureau and designed to show how communities are changing. Through asking questions of a sample of the population, it produces national data on more than 35 categories of information, such as education, income, housing, and employment.

During the Activity

Teachers could direct students to work in groups of two, three, or four to encourage collaboration and discussion.

After the Activity

Teachers will facilitate a class discussion in which students share and respond to each other's observations and arguments about the data. Teachers should tell students to use these sentence starters:

- I notice …

- I wonder …

- Similarities I see are …

- Differences I see are …

- It is surprising that …

Extension Ideas

- Teachers could ask students to gather data from *data.census.gov* on the median earnings in their state for men and women whose highest levels of educational attainment are either a high school diploma (or equivalent) or a bachelor's degree. Students could then compare data for their state with data for the nation (from **Item 1**).

- Teachers could adapt this activity for different GAISE levels:

  - GAISE Level B: Teachers could facilitate students' collection of external data on sex and GPA. Students would create two box plots — one for men and one for women — and could compare information from the two box plots to draw conclusions about the data.

  - GAISE Level C: Teachers could facilitate students' collection of data on sex and a quantitative variable of their choice. Students would create two box plots — one for men and one for women — and could compare information from the two box plots to draw conclusions about the data and identify any potential sampling error due to their methodology.

Student Activity

Click *here* to download a printable version for students.

Activity Item

The following item is part of this activity and appears at the end of this student version.

- Item 1: Box Plots

Student Learning Objectives

- I will be able to interpret a box plot.
- I will be able to compare different variables across data sets.
- I will be able to compare box plots to make claims and inferences based on the data.

Item 1: Box Plots shows national median earnings — for all 50 states and the District of Columbia — for men and women aged 25 and older whose highest levels of educational attainment are either a high school diploma (or equivalent) or a bachelor's degree.

Inspect the two sets of box plots. Now use them to answer the following questions.

1. Compare the median earnings for people whose highest level of education is a high school diploma or equivalent with those for people whose highest level of education is a bachelor's degree. What observations can you make by just looking at the data? Justify your claim.

   Student answers will vary but could include: People whose highest level of education is a high school diploma or equivalent typically earn less than those whose highest level of education is a bachelor's degree, based on the difference in median earnings between the two levels of education.

2. Compare the median earnings of men and women. What observations can you make by just looking at the data? Justify your claim.

   Student answers will vary but could include: Men typically earn more than women, based on the difference in median earnings between the sexes (male and female).

3. Estimate the values for each of the measures (minimum, first quartile, etc.) in **Item 1** and show your work below for men whose highest level of education is a high school diploma or equivalent, using the relevant box plot.

   - **Minimum: Actual value is $28,114; student answers should be between $27,000 and $29,000.**

   - **First quartile: Actual value is $31,429; student answers should be between $30,500 and $32,500.**

   - **Median: Actual value is $33,252; student answers should be between $32,000 and $34,000.**

- **Third quartile: Actual value is $36,091; student answers should be between $35,000 and $37,000.**

- **Maximum: Actual value is $42,791; student answers should be between $42,000 and $44,000.**

- **Range: Actual value is $14,677; student answers should be between $14,000 and $16,000.**

- **Interquartile range: Actual value is $4,662; student answers should be between $4,000 and $5,000.**

4. Estimate the values for each of the measures for women whose highest level of education is a high school diploma or equivalent, using the relevant box plot.

- **Minimum: Actual value is $18,834; student answers should be between $18,000 and $20,000.**

- **First quartile: Actual value is $20,903; student answers should be between $20,000 and $22,000.**

- **Median: Actual value is $21,662; student answers should be between $21,000 and $23,000.**

- **Third quartile: Actual value is $24,177; student answers should be between $23,000 and $25,000.**

- **Maximum: Actual value is $28,484; student answers should be between $27,000 and $29,000.**

- **Range: Actual value is $9,650; student answers should be between $8,000 and $10,000.**

- **Interquartile range: Actual value is $3,274; student answers should be between $3,000 and $4,000.**

5. Estimate the values for each of the measures for men whose highest level of education is a bachelor's degree, using the relevant box plot.

- **Minimum: Actual value is $45,100; student answers should be between $44,000 and $46,000.**

- **First quartile: Actual value is $53,445; student answers should be between $52,000 and $54,000.**

- **Median: Actual value is $56,516; student answers should be between $55,000 and $57,000.**

- **Third quartile: Actual value is $63,088; student answers should be between $62,000 and $64,000.**

- **Maximum: Actual value is $74,968; student answers should be between $74,000 and $76,000.**

- **Range: Actual value is $29,868; student answers should be between $29,000 and $31,000.**

- **Interquartile range: Actual value is $9,643; student answers should be between $9,000 and $11,000.**

6. Estimate the values for each of the measures for women whose highest level of education is a bachelor's degree, using the relevant box plot.

- **Minimum: Actual value is $31,246; student answers should be between $30,000 and $32,000.**

- **First quartile: Actual value is $36,599; student answers should be between $36,000 and $38,000.**

- **Median: Actual value is $39,129; student answers should be between $38,000 and $40,000.**
- **Third quartile: Actual value is $42,551; student answers should be between $42,000 and $44,000.**
- **Maximum: Actual value is $57,663; student answers should be between $56,500 and $58,500.**
- **Range: Actual value is $26,417; student answers should be between $26,000 and $28,000.**
- **Interquartile range: Actual value is $5,952; student answers should be between $5,000 and $7,000.**

7. Compare the minimum median earnings for men whose highest level of education is a bachelor's degree with the third quartile of median earnings for women whose highest level of education is a bachelor's degree. Explain what this means in context.

   Student answers will vary but could include: The minimum median earnings for men whose highest level of education is a bachelor's degree is about $45,000. The third quartile of median earnings for women whose highest level of education is a bachelor's degree is less than $45,000. In the context of the data, this means that the median earnings for men at this education level are greater than the median earnings for women at this education level in at least 75 percent of the states.

8. Compare the third quartile of median earnings for men whose highest level of education is a high school diploma or equivalent with the first quartile of median earnings for women whose highest level of education is a bachelor's degree. Explain what this means in context.

   The third quartile of median earnings for men whose highest level of education is a high school diploma or equivalent and the first quartile of median earnings for women whose highest level of education is a bachelor's degree are both about $36,000. In the context of the data, this means that the median earnings for men at the high school diploma level of educational attainment are greater than $36,000 in about 25 percent of the states, whereas the median earnings for women at the bachelor's degree level of educational attainment are less than $36,000 in about 25 percent of the states.

9. Compared with the 50 states, the District of Columbia is an outlier in this data set in terms of median earnings for women whose highest level of education is a bachelor's degree (represented by the second * in Item 1). Examine the relevant box plot to answer the following questions about outliers:

   a. Estimate the value of this outlier.

      Actual value is $57,663; student answers should be between $55,000 and $60,000.

   b. For a value to be considered an outlier, it must be greater than (1.5*IQR) + Q_3 or less than Q_1 – (1.5*IQR). Verify below, showing your work, that your estimate from question 9a would qualify as an outlier (use estimates of quartiles).

      While students will use estimates, the actual computed value is:

      (1.5*IQR) + Q_3

      (1.5*$5,952) + $42,551 = $51,479

$57,663 > $51,479

Therefore, $57,663 is an outlier.

c. Why do you think the median earnings for women in the District of Columbia whose highest level of education is a bachelor's degree are noticeably greater than the median earnings for women of the same education level in the nation as a whole?

The median earnings for those women in the District of Columbia may be significantly greater than the median earnings for those women in the nation as a whole because the District of Columbia, is an entirely urban area, which tends to have a higher cost of living than other areas, while the rest of the nation contains both urban and nonurban areas. It could also be because federal government jobs in the nation's capital may pay a higher salary.

(At this point in the activity, teachers could encourage a group brainstorm and assessment of possible reasons for this data observation.)

d. Describe the impact of this outlier on the mean, median, and shape of the distribution.

This outlier significantly increases the mean, does not have a significant impact on the median, and causes the shape of the distribution to skew right.

10. At which level of educational attainment do the median earnings have the most variability? Explain.

There is greater variability in median earnings for people whose highest level of education is a bachelor's degree compared with those whose highest level of education is a high school diploma or equivalent. The range and interquartile range are greater for people at the bachelor's degree level, as is apparent in the comparatively elongated shape of the box plot.

(At this point in the activity, teachers could prompt struggling students to connect a measure of variability they know, such as the range or IQR, with the same feature in a box plot.)

11. Compare box plots to assess the differences in median earnings at each level of educational attainment (use estimates of quartiles).

a. How much greater are the median earnings of men whose highest level of education is a bachelor's degree than those of men whose highest level of education is a high school diploma or equivalent?

Actual computed value is $23,264; student answers should be between $22,000 and $24,000.

b. Approximately how much more would a man whose highest level of education is a bachelor's degree earn over a 40-year career compared with a man whose highest level of education is a high school diploma or equivalent?

Actual computed value is $930,560 ($23,264 ∗ 40); student answers should be between $880,000 and $960,000.

   c. How much greater are the median earnings of women whose highest level of education is a bachelor's degree than those of women whose highest level of education is a high school diploma or equivalent?

Actual computed value is $17,467; student answers should be between $16,000 and $19,000.

   d. Approximately how much more would a woman whose highest level of education is a bachelor's degree earn over a 40-year career compared with a woman whose highest level of education is a high school diploma or equivalent?

Actual computed value is $698,680 ($17,467 ∗ 40); student answers should be between $640,000 and $780,000.

12. Compare box plots to assess the differences between the sexes in terms of median earnings (use estimates of quartiles).

   a. Among people whose highest level of education is a high school diploma or equivalent, about how much greater are the median earnings for men than those for women?

Actual computed value is $11,590; student answers should be between $10,000 and $13,000.

   b. Among people whose highest level of education is a bachelor's degree, about how much greater are the median earnings for men than those for women?

Actual computed value is $17,387; student answers should be between $16,000 and $19,000.

   c. Why do you think women have lower median earnings than men in both of these cases? Keep in mind that the data only relate to women who are in the workforce and exclude homemakers and retirees.

Student answers will vary but could include: Women may have lower median earnings than men in these cases because of various factors, including occupational choice and possible wage discrimination.

13. Compare the gaps in national median earnings between the sexes and between the levels of educational attainment (use estimates of quartiles).

   a. In just looking at the data in the box plots, which appears to be greater — the difference in earnings between the sexes or the difference in earnings between the levels of educational attainment? Explain.

Student answers will vary, but students should use appropriate reasoning to justify their answer.

   b. Use math to determine whether your prediction from question 13a is correct or incorrect (use estimates of quartiles). Explain.

Student answers will vary depending on their choice in 13a. The calculations and explanations for the correct prediction (education level) could include:

Looking at the data for women in both categories, the difference between the median of the median earnings for women whose highest level of education is a bachelor's degree (about $40,000; actual value: $39,129) and that for women whose highest level of education is

a high school diploma or equivalent ($21,000; actual value: $21,662) is about $19,000 (actual value: $17,467).

Looking at the data for men in both categories, the difference between the median of the median earnings for men whose highest level of education is a bachelor's degree ($57,000; actual value: $56,516) and that for men whose highest level of education is a high school diploma or equivalent ($33,000; actual value: $33,252) is about $24,000 (actual value: $23,264).

Looking at the bachelor's degree education level data, the difference between the medians of the median earnings for men ($57,000; actual value: $56,516) and for women ($40,000; actual value: $39,129) is about $17,000 (actual value: $17,387). For the high-school-diploma education level data, the difference between the medians of the median earnings for men ($33,000; actual value: $33,252) and for women ($21,000; actual value: $21,662) is about $12,000 (actual value: $11,590).

Because the median earnings differences between the sexes within a particular education level ($17,387 and $11,590) are less than the median earnings differences between highest education levels for either men ($23,264) or women ($17,467), the education level gap is larger.

14. Evaluate a box plot as a way of displaying a data set.

   a. What are the advantages of box plots?

    Student answers will vary but could include: Box plots provide a quick visualization of the 5-number summary (minimum, first quartile, median, third quartile, and maximum) and help show the center, spread, and outliers of a data set. Box plots are effective for comparing different data sets as well.

    b. What are the limitations of box plots?

    Student answers will vary but could include: Box plots do not show exact values and, compared with some other graphs, their data are not as easy to digest quickly. Also, box plots do not provide a quick visualization of the mean.
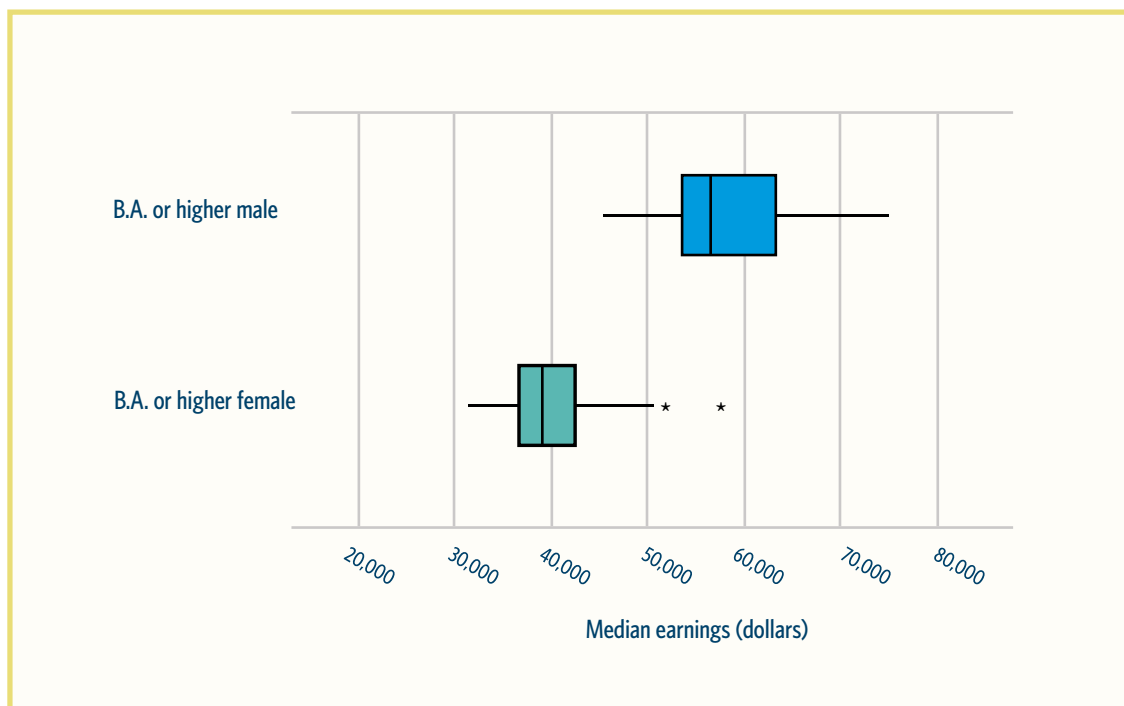
Item 1: Box Plots

Distribution of U.S. State Median Earnings for Men and Women Whose Highest Level of Education Is a High School Diploma or Equivalent
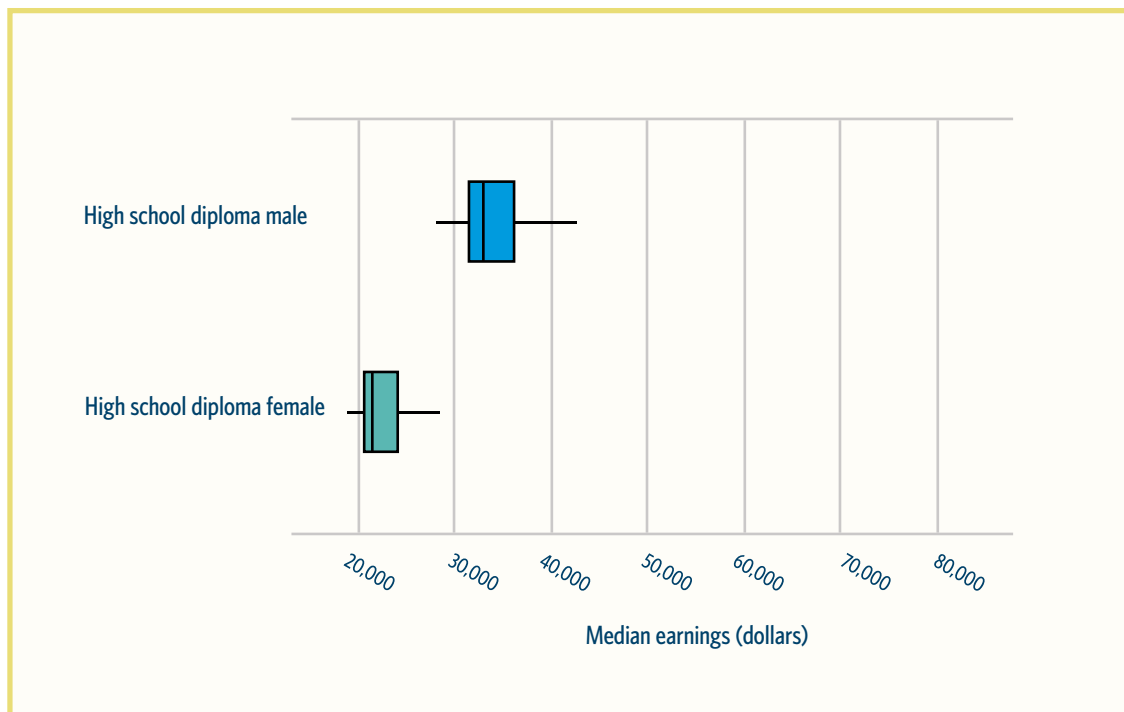
Item 1: Box Plots (Continued)

Distribution of U.S. State Median Earnings for Men and Women Whose Highest Level of Education Is a Bachelor's Degree

Source for data: U.S. Census Bureau, Median Earnings in the Past 12 Months (in 2013 Inflation-Adjusted Dollars) by Sex by Educational Attainment for the Population 25 Years and Over.
2009-2013 American Community Survey 5-Year Estimates.